基于用户分群的数字社区消费者多模态特征分析与服务效能提升研究

黎灿垚[1]，韦　伟[1]，刘晓丽[1]，周林兴[2*]，王　帅[2]

（1. 广西中烟工业有限责任公司，南宁　530001；2. 上海大学文化遗产与信息管理学院，上海　200444）

摘　要：[目的 / 意义]对数字社区消费者进行多模态特征分析与服务效能提升，有助于为数智赋能在线社区建设提供新视野、为相关部门部署数字决策提供新动能。[方法 / 过程]结合社区特性构建用于消费者分群的数据维度，将维度下的 24 个指标数据进行二次聚合后实现分群，并构造参数、决策变量及函数表，从而分析消费者多模态特征，基于这些特征实现数字消费服务效能的提升。[结果 / 结论]实证分析结果表明，本文模型能够生成合理有效的分群结果，进而实现类群特征区分以及群间渗透与漂移现象分析；分群结果呈现出 6 类消费者群体：重点、中心、特殊、沉睡、流失和一般类群，绝大多数类群都会产生用户渗透现象，仅有一般用户类群会发生群间漂移现象；服务效能提升模型表明最受关注价值的群体为中心和重点类群。

关键词：用户分群；AP-DBSCAN；多模态特征；数字社区；数字消费

中图分类号：G250　　文献标识码：A　　文章编号：1002-1248（2023）02-0030-15

引用本文：黎灿垚，韦伟，刘晓丽，等. 基于用户分群的数字社区消费者多模态特征分析与服务效能提升研究[J]. 农业图书情报学报，2023，35(2)：30-44.

1 引　言

近年数字消费为居民打开了智能生活新格局，其定义是通过数字社区对产品服务进行推广及销售的实践活动[1]。“十四五”规划纲要明确提出数字中国战略，旨在把握时代机遇、加快数字经济发展，以数字化转型催生产业新模态、壮大经济新引擎[2]，数字消费

收稿日期：2023-01-06

基金项目：广西中烟工业有限责任公司科技项目“基于机器学习方法的营销活动效果动态评估”（CGAXZX20210030050001-044）；江苏省社会科学基金青年基金“社会感知数据驱动下的公共卫生事件时空演化研判机制研究”（20TQC001）；中国博士后科学基金特别资助“面向应急管理的时空数据语义模型构建及创新应用机理研究”（2021T140311）；中国博士后科学基金面上项目“环境污染突发事件的时空数据挖掘及协同治理机制研究”（2019M650108）

作者简介：黎灿垚（1972- ），男，硕士研究生，研究方向为卷烟互联网营销工作研究与实施管理。韦伟（1986- ），女，研究方向为卷烟行业互联网营销及研究。刘晓丽（1978- ），女，研究方向为卷烟行业互联网营销及研究。王帅（1996- ），男，博士研究生，研究方向为数据分析与挖掘

*通信作者：周林兴（1974- ），男，教授，博士生导师，研究方向为政府数据治理。Email：sobzyl@hhu.edu.cn

成为当今创新领跑的重要生产力。面向数字社区呈现出的新特征进行研究，有助于把握消费者行为风向和心理变化，使企业和相关单位更充分地融入数字化队列。然而，数字消费活动重视渠道成本、泛化自媒体宣传和忽视社群维系的发展路径，同样为消费者带来产品内容同质化、信息虚假化和零售模式单一化等不可忽视的负面影响[3]。受同类问题影响，消费者表现出一定的潜在抗拒态势，具体而言集中在群体个性化不突出、内容接受层次浅和资产意识缺失方面。究其原因，数字产品供应厂商并未真正理解消费群体，缺乏与偏好高度匹配的精细运营模式，使其较难通过数字社区消费活动满足实际需求，如何以消费者为中心开展服务优化逐渐成为研究热点。消费者是数字社区活动的参与主体，也是数字信息消费主体[4]，基于分群技术对其进行深入分析和特征解读能够保障主体满意度，进而通过服务布局优化提升社区运营效能。基于此，本研究面向数字社区消费者进行分群，并开展社区服务效能提升分析，为数字中国建设提供新视野、为数智赋能提供新方案。

2　研究现状

2.1　数字社区消费活动类型与内容

数字社区消费活动可以归结为 4 种类型：分别是区域、媒介、产品和新技术类。区域类划分依赖消费者居住地，横向上分为境内[5]和境外[6]，纵向上通过中国行政区域管辖范围确定。媒介类以活动传播载体为划分依据，包括软件和硬件两种，对于前者，又涉及基于“互联网＋平台”[7]进行的、基于 APP 或社群[8]和基于生态圈[9]运作的 3 类情形；后者则涵盖 VR 技术下的可穿戴设备[10]和新型智能终端[11]两种物理载体。产品类消费活动以商品或品牌宣传为目标导向铺设数字化方案，以产品功能实现[12]、信息集成[13]和数字资产[14]等内容为基点开展推广活动。数字类特指运用 AI、AR 和眼动追踪等高精尖技术，针对某一品牌进行战略创新而提出 SEO 或 PPC 等数字消费推广策略进而建立生态信息反馈渠道的方案[15]。数字社区研究内容主要归纳为 4 个方面：活力激发与运营策略[16]，重点对数字活动运营机制进行分析精准化内容传播与用户解读[17]，重点对精细化数字内容推荐机制和消费用户进行分析；人才培育及能力建设[18]，重点对国家数字化战略下的人才能力培育策略进行分析；前沿理论与技术应用[19,20]，重点分析领域前沿理论和技术的应用场景。

2.2　用户画像与分群研究

用户画像起源于电商领域，其定义包括 3 个方面内容[21]，分别是作为画像构建前提的数据搜集工作、作为体现画像强关联性的业务关联工作和作为用户解读基础的数据挖掘工作。画像目的是把用户抽象成差异性标签进行描述，从而实现特征区分，对用户开展行为、需求和偏好解读[22]。在画像实现方面，现有研究主要可以分为 4 类，分别是面向用户偏好的实现方案[23]、面向情绪的实现方案[24]、面向主题的实现方案[25]和面向用户行为的实现方案[26]。能够发现，画像技术已经在图情领域广泛应用[27]，与其较为接近的技术是用户分群。用户分群是从用户画像概念下衍生出来的二次聚合概念，延伸到消费活动方面，分群对企业来说是一种降低成本且提高用户解析效率的方法[28]，相较于画像，它能够归结群体特点、群体网络、交互关系和行为偏好等多模态信息。目前，用户分群方法使用率较低，较有代表性的是占张帆[29]面向产品知识推送所开展的用户分群研究，该研究在分析用户偏好基础上，对分群模型进行形式化定义并利用偏好扩散和提取两种特征基于聚类算法设计分群方案。实现技术主要分为数学建模和机器学习，前者指通过设计专门算法实现分群，如宋嵩[30]面向 URL 数据设计专门算法实现分群；后者指将既有算法输入到机器学习模型中进行训练实现自动分群，如吴艳玲和孙思阳[31]采用自适应自然梯度算法实现用户聚类的最佳分群效果。

2.3　在线服务效能提升研究

在线服务作为移动互联网时代新的服务方式，与其他相对成熟的模式相比在交互层次、支撑模式与体

验感知方面存在巨大差异，特别是在线消费活动常伴随大量的即时通讯消息和互动[32]，如何保证在线服务质量和水平、提升用户感知满意度已成为亟待解决的问题。现有研究可以根据目标导向分为 5 类，分别是：以管理效能提升为导向，如陈长庆[33]以途牛网为例研究服务质量改进对平台盈利的影响；以服务效能优化为导向，如王婉等[34]运用扎根理论分析用户访谈质性资料，并构建数智环境下在线社区用户服务需求的理论模型；以环境效能改善为导向，如张邦辉等[35]发现在线政务服务为营商环境带来了“数字红利”，可以通过赋能在线服务环境提高企业用户营商评价和市场化程度；以布局调整为导向，如于丽娟[36]通过分析“互联网 + 智慧服务”优化数字阅读精准推广模式的意义，从多维度剖析优化布局下的创新机制和发展驱动力；以用户价值为导向，如王毅和吴睿青[37]结合画像提出公共图书馆用户价值优化策略。

可以发现，既有研究虽然覆盖有相当一部分针对用户开展的分析，但这些分析主要集中在精细化运营方案的制定方面，尚缺乏数字社区服务视角下的效能提升研究。另一方面，画像技术虽然能够较贴切地呈现用户信息概貌，但群间多模态特征的横向比对和差异性归因能力一般。基于此，本文从价值发现角度，通过聚合画像实现分群，据此分析数字社区消费者群体的多模态特征，并提出服务效能提升方案。

3 研究设计

3.1 面向数字社区消费者的多模态特征分析

3.1.1 分群指标体系建立

多模态特征分析，是指对多源异构数据融合与特征挖掘的过程。经对实证数据进行融合与特征挖掘，共提取到画像、群偏好、活动影响力和群消费能级 4 个特征数据维度。为便于表述将数字消费活动简称为“活动”、消费者简称为“用户”。

（1）画像维度。观察国内外用户画像构建特点，发现所用指标可分为用户基本信息[38]、社交情况[39]和 UGC[40]（Users Generated Contents，UGC）三大类，画像指标及对应含义如表 1 所示。

表 1 中的用户地区、UGC 主题和关键词则须经数据加工过程才能转化为画像标签和分群依据，过程如下：用户地区按省级（或直辖）管辖区域分为华北地区、华东地区、东北地区、华中地区、华南地区、西南地区、西北和港澳台及海外地区；UGC 主题、关键词使用 LDA 进行提取和挖掘；情感倾向抽取参照情感识别方法[24]进行；将活动时长和社交指标下的点赞、评论情况按极值区间均等化分为低、中、高程度。

（2）群偏好维度及数据加工。挖掘群偏好有助于发现用户行为背后的潜在知识和问题[41]，为有效反映

表 1　用户画像指标

Table 1　User profile indicators

一级指标	二级指标	对应含义
用户基本信息	用户昵称	参与活动用户的 Users ID
	用户性别	参与活动用户的性别
	用户年龄	参与活动用户的年龄
	用户地区	参与活动用户所在地区
	活动时长	用户参与活动的时间跨度（单位：d）
社交情况	文本点赞数	发帖点赞与回帖点赞数之和
	文本评论数	发帖评论与回帖评论数之和
UGC 内容	UGC 主题	用户生成的文本内容主题
	UGC 关键词	用户生成文本内容主题下的关键词
	UGC 情感倾向	用户生成 UGC 的情感倾向

数字社区用户原貌、提升分群科学性，筛选出用于分析的 4 个一级指标，如表 2 所示。

表 2 需要进行数据加工的二级指标有 5 个：其中活动、消费频率相对指数和活动倾向评分利用力度指数公式实现，据值域分为“低、中、高”3 个层次；产品倾向通过 Logistic 回归实现；兴趣矩阵量级参照势能公式改写计算，包括“低、中、高”3 个层次；活动参与时间按早（02:00—10:00）、午（10:01—18:00）、晚（18:01—02:00）划分。

（3）群活动影响力维度。包括 3 个一级指标，分别是：①平均用户价值计算得分，用于评估用户对活动主体的整体价值，通过访客平均（UV）价值公式计算；②用户裂变病毒系数 K 值，用于表现用户裂变能力并侧面表征其潜在价值，通过病毒系数计算（访客邀请—转化率的百分比形式）；③用户盈利能力计算得分，用于反映数字社区创收价值，据经济学获利能力模型算得分值并按值域均分为“低、中、高”3 个层级输出。

（4）群消费能级维度。包括两个方面的内容，分别是商业价值评定[42]和消费倾向预测[43]，将其作为维度下的一级指标使用。前者包括自适应决策边界下的消费层级、数值计算下的消费概率和 GBDT 下的消费能力；后者则包括基于线性回归的显隐性商业价值水平评定两个角度；由此生成 5 个二级指标并按均等值域区间将上述内容划分为“低、中、高”三档。

3.1.2　分群过程及其实现

实现分群的第一步是用户分层，即将用户按各自特征分成若干个互不重叠的部分，每一部分称之为“层”，目的在于对用户进行初步分类[44]进而反映其需求，但分层方法易忽视行为数据落差、割裂用户间交互情况，难从全局解读用户参与数字社区活动所表征现象的动因，因此本研究通过画像构建达到同等分层效果。分层后对画像数据进行聚合从而实现分群，并按 AARRR（又称“海盗模型”，Acquisition、Activation、Retention、Revenue、Referral）模型[45]改写后进行命名，分群过程设计如图 1 所示。

图 1 所示分群过程如下：第一步，对数据集进行预处理；第二步，将预处理后的用户原始数据映射到指标体系的画像、偏好、活动和消费 4 个维度中，并利用 visualMap 实现从指标体系到标签关联数据集的构建；第三步，使用 AP（Affinity Propagation，AP）算法对标签数据聚合实现画像，该算法基于数据点间“信息传递”的聚类方式[46]通过吸引度和归属度计算各节点消息传播强度来寻取聚类中心；第四步，将画像数据组合成画像库数据集，通过 DBSCAN（Density-Based Spatial Clustering of Applications with Noise，DBSCAN）聚合数据实现分群；第五步，对用户类群下的数据流量进行变分，结合用户群特点和 A（Acquisition）-A（Activation）-R（Retention）-R（Revenue）-R（Refer）模型实现群体命名，特别地，数字社区消费用

表 2　群偏好维度对应指标

Table 2 Corresponding indicators of group preference dimensions

一级指标	二级指标	含义
行为偏好	活动参与时间	描述用户参与活动的时间概要
	活动频率相对指数	描述单一用户较全体而言的活动参与程度
消费偏好	消费频次	描述用户活动消费次数
	消费频率相对指数	描述单一用户较全体而言的消费概率程度
	产品倾向	描述用户对活动相关的产品倾向
兴趣偏好	兴趣矩阵量级	描述用户对活动的整体兴趣
	全局兴趣偏好	描述用户在活动全局最关注的兴趣点
活动类型偏好	活动倾向评分	描述用户对活动形式好感度
	主要活动倾向	描述用户对活动形式的偏好倾向

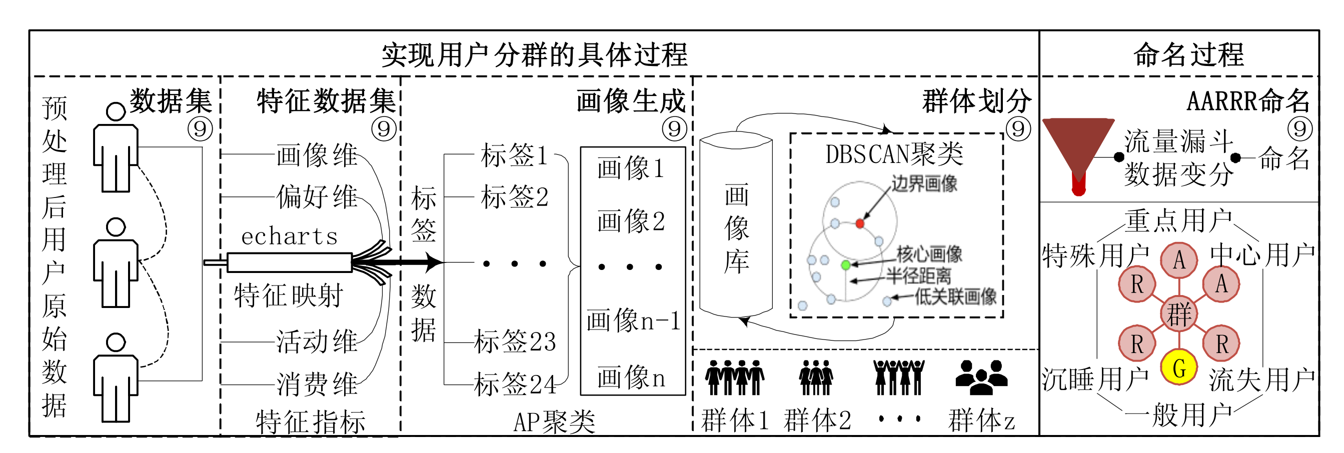


图 1　分群过程

Fig.1 Clustering process

户基数最大的群体应当是一般类群，因此在 AARRR 模型中额外补充关于“G（General）”的内容。

3.1.3　群交互网络关系分析

社会网络分析（Social Network Analysis，SNA）常用于研究社会个体成员之间存在的关系及其对社会网络整体结构或内部个体带来的影响[47]，是一种定量的行动者交互行为研究方法。本研究利用胡昌平等[48]提出的群体交互判别方法进行网络构造和关系解读：从用户群共现角度出发通过群体数据构建标签，将具有潜在关联的标签数据链接两个不同的群体进而组成网络。在关系解读方面，将关联标签数量作为共现关系强度度量方式，强度越大表明相邻两个群之间的关联越密切；利用中心性解释不同群体对活动资源占有量的大小。

3.1.4　群间渗透与漂移现象

用户群间渗透指某群体中部分用户趋向其他类群的情形，漂移指某群体内所有用户分别趋向其他类群、使原有类群逐渐消解却又不产生新类群的情形。用户因自身利益趋向，极有可能在群间出现渗透和漂移现象，发生机理详见图 2。

图 2 渗透示意图表明，用户群中含有一部分趋向用户，当其参与活动后产生离群行为，离群后的用户不断涌入其他类群中，削弱原本群内用户基数并壮大其余类群力量。对于渗透现象的识别，主要依赖分支定界法。当群内用户均具有离群趋向时，会在参与活动后定向漂移到新类群中，与渗透不同的是用户所属

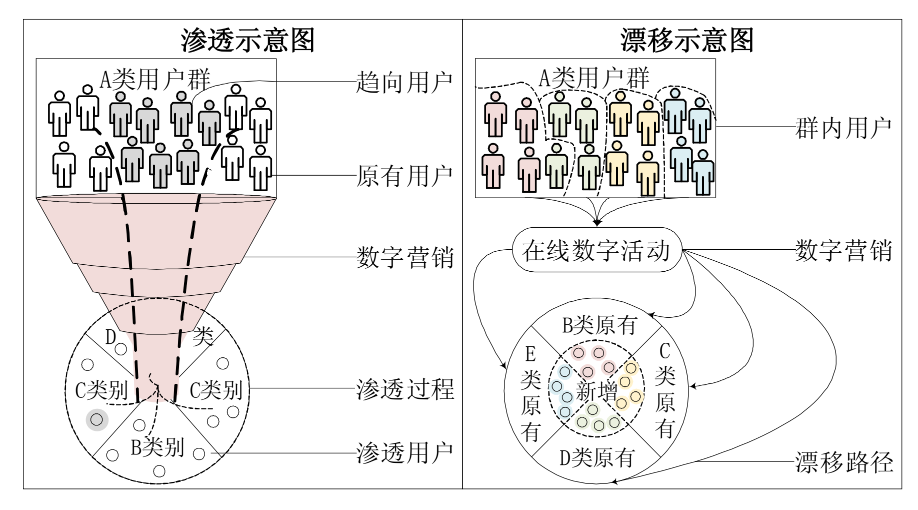


图 2　群间渗透与漂移现象发生机理

Fig.2 Occurrence mechanism of inter-group infiltration and drift

原类群不复存在并将转化为其他类群的新增用户。对于漂移用户的识别，主要基于开源求解器 SCIP 中的 Relpscost 算法[49]完成。

3.2　数字社区消费服务效能提升

本研究参照孟秀丽等[50]提出的方法，结合用户分群结果和群间特性确定服务效能提升模型并得出用户最优行为及均衡条件，探讨数字社区服务对用户的影响，从而提升数字社区活动质量。

3.2.1　参数及决策变量函数建立

数字消费是由供应方、消费者和数字社区组成的三层网络结构，其原理包括两条主线，分别是信息流与商品流。信息流是由供应方 α 在考虑承担的成本和服务质量后将需求发送给数字社区 β，消费者根据自身因素决定是否参与活动。商品流是指消费者 γ 消耗一定时间或其他成本换取与活动相关的服务或产品。相较于供应方，消费者和数字社区都可以通过活动受益，因此将其二者受益比例设置为 ε，供应方的受益比例为 $1-\varepsilon$，则供应方可以通过数字社区中的消费者感知获益对自身及三方服务质量进行控制。数字社区关于消费者感知受益所需承担的受益损失比例为 μ，则供应方关于自身感知受益所实际付出的时间、精力或其他成本比例为 $1-\mu$，也就是供应方可以通过控制消费者的感知受益来保障服务质量。据此构建的参数及决策变量函数如表 3 所示。

活动服务感知受益和实际活动受益满足约束条件，如式（1）：

$$0 \leqslant \hat{s}^t < 1;\ 0 \leqslant s^t < 1 \tag{1}$$

其中，$\hat{s}^t$ 和 s^t 越大代表活动服务质量水平越高，1 则代表完美水平，0 代表极差水平。t 时间段内供应方对活动的感知服务质量为 $\hat{s}^t$，受其他活动或不同时期同一活动的实际质量 s^{t-1} 与感知质量 $\hat{s}^{t-1}$ 的影响，参考 NAGURNEY 等[51]所作的研究，假设负面影响因素出现的概率 P_j 是活动平均感知服务质量 $\bar{s}$ 和实际质量 s^t 的函数，那么活动平均感知质量为 $\bar{s}=1/n(\sum_{\beta=1}^{n}\bar{s}^t)$。上述内容以服务质量为通道将供应方、消费者和数字社区连接起来，对三层网络结构相关的参数、决策变量及函数进行描述，有助于构建效能提升模型并提出针对性策略。

表 3　参数及决策变量函数表

Table 3 Functions of parameters and decision variables

参数表		决策变量与函数表	
符号	说明	符号	说明
t	消费者活动周期/活动次数 t=1,2,…,T	$q^t_{\gamma\beta}$	t 内 α 与 β 间成本量，构成 $p^t_2 \in R_+^{no}$
γ	活动参与消费数 γ=1,2,…,m	$q^t_{\gamma\beta}$	t 内 α 与 β 间成本衰减比例为 $p^t_3 \in R_+^{no}$
β	活动涉及数字社区平台数 β=1,2,…,n	$\hat{q}^t_{\gamma\beta}$	t 内 α 承受的成本提升率为 $p^t_4 \in R_+^{o}$
α	供应方在同一周期内活动发起数 α=1,2,…,o	$q^{rt}_{\alpha\beta}$	t 内活动感知服务质量为 $p^t_5 \in R_+^{o}$
φ	供应方在活动中的总负担成本比例	p^t_k	t 内活动实际服务质量为 $s_1 \in R_+^{o}$
λ_γ	消费者的服务质量成本系数	p^{rt}_k	t 内消费者流失率 $s_2 \in R_+^{n}$
λ_β	数字社区的服务质量成本系数	$\hat{s}^t$	t 内数字社区的运营成本
ρ	群体对实际活动传播效果的敏感系数	P_β	t 内供应方与消费者间成本差值
ε	数字社区在活动中的受益比例	$f_\beta(Q_1^t)$	t 内数字社区与供应方成本差值
μ	数字社区在活动中消耗的成本比例	$\hat{f}(Q_2^t)$	t 内数字社区提升感知受益的成本
ω	消费者实际受益系数	$c_{\gamma\beta}(Q_1^t)$	t 内 α 负担的数字社区平台运营成本
$\hat{q}^t_{\alpha\beta}$	t 内 γ 与 β 间交互次数，构成 $Q^t_1 \in R_+^{mn}$	$\hat{c}_{\alpha\beta}(q_2^t,s_2)$	t 内消费者的受益衰减幅度
$q^t_{\gamma\beta}$	t 内 α 与 β 间交互次数，构成 $Q^t_2 \in R_+^{no}$	$c_\alpha(Q_2^t)$	t 内 α 关于提升消费者感知受益的成本
$q^t_{\gamma\beta}$	t 内 α 与 β 间有效交互数构成 $Q^t_3 \in R_+^{no}$	$c^r_\alpha(Q_3^t)$	t 内消费者的活动感知受益需求
$q^t_{\gamma\beta}$	t 内 γ 与 β 平均消费价格构成 $p^t_1 \in R_+^{mn}$	$d^t_\alpha(p_4^t,s_1)$	t 内消费者的受益衰减承受程度

3.2.2 服务效能提升模型构建

为建模需要作如下假设：供应方提起的活动在一定时间内会由消费者加入；消费者所感知到的活动服务质量是由供应方和数字社区共同作用的结果；与已有关于网络均衡的假设一致，供应方、数字社区和消费者均为理性决策者，且以自身利益最大化为目标，三方网络成员通过 Nash 非合作竞争达到均衡博弈状态；供应方、数字社区与消费者的成本函数是连续可微凸函数。据此构建服务效能提升的概念模型内容：供应方、数字社区和消费者的最优行为及均衡条件分析，用于提出三层网络结构主体的优化方案；模型求解，用于求得模型数值、通过数值分析逆向改善服务质量；将数值优化对应的指标纳入到效能提升树，从可视化构建出发，呈现活动所处不同阶段应采取的具体策略。

最优行为及均衡条件求解：消费者位于信息流末端、商品流第一层，为使服务质量提升取得更好效果，优先考虑其最优行为，其目标函数如式（2）所示：

$$\max H_{\beta}=\sum_{t=1}^{T}\sum_{k=1}^{o}\hat{p}_{\alpha\beta}^{t}\hat{q}_{\alpha\beta}^{t}+\sum_{t=2}^{T}\sum_{k=1}^{o}p_{\alpha\beta}^{rt}q_{\alpha\beta}^{rt}\varepsilon-\sum_{t=1}^{T}c^{t}{}_{\gamma\beta}(q_{\gamma\beta}^{t})-\sum_{t=1}^{T}\hat{c}_{\alpha\beta}^{t}(\hat{q}_{\alpha\beta}^{t},s^{t})-\sum_{t=1}^{T}h_{\beta}(s^{t})q_{\gamma\beta}^{t}-q_{\alpha\beta}^{r}\omega p_{\beta}\vartheta \quad (2)$$

它表示消费者的感知受益为数字社区活动传播效果和供应方负担成本减去自身感知亏损的整体感知受益水平，其中 $h_{\gamma}(s^{t})=1/2\lambda_{\gamma}(s^{t})^{2}$ 为数字社区宣传成本。当 $(Q_{1}^{t^{*}},Q_{3}^{t^{*}},s_{2}^{*})\in\Phi$ 时满足$(s^{t}-s^{t^{*}})\geq 0;q_{\gamma\beta}^{t^{*}}>0,p_{\gamma\beta}^{t^{*}}=\partial f_{\gamma}^{t}(Q_{1}^{t})/\partial q_{\gamma\beta}^{t}$，即对消费者而言，均衡状态下的活动发起成本与数字社区宣传成本损失之和等同于受益水平。同理，对于数字社区和供应方，其目标函数分别如式（3）和式（4）所示：

$$\max H_{\beta}=\sum_{t=1}^{T}\sum_{k=1}^{o}\hat{p}_{\alpha\beta}^{t}\hat{q}_{\alpha\beta}^{t}+\sum_{t=2}^{T}\sum_{k=1}^{o}p_{\alpha\beta}^{rt}q_{\alpha\beta}^{rt}\varepsilon-\sum_{t=1}^{T}c^{t}{}_{\gamma\beta}(q_{\gamma\beta}^{t})-\sum_{t=1}^{T}\hat{c}_{\alpha\beta}^{t}(\hat{q}_{\alpha\beta}^{t},s^{t})-\sum_{t=1}^{T}h_{\beta}(s^{t})q_{\gamma\beta}^{t}-q_{\alpha\beta}^{r}\omega p_{\beta}\vartheta \quad (3)$$

$$p_{\alpha\beta}^{rt}+c_{\alpha}^{rt}\begin{cases}=p_{\alpha}^{rt},q_{\alpha\beta}^{rt}>0\\ \leq p_{\alpha}^{rt},q_{\alpha\beta}^{rt}=0\end{cases};d_{\alpha}^{rt}\begin{cases}\leq\sum_{\alpha=1}^{o}q_{\alpha\beta}^{rt},p_{\alpha}^{rt}=0\\ =\sum_{\alpha=1}^{o}q_{\alpha\beta}^{rt},p_{\alpha}^{rt}>0\end{cases} \quad (4)$$

式（3）表示数字社区的自身获益为部分供应方担负成本和消费者感知亏损之和，其中 为数字社区进行活动宣传的实际成本；式（4）可以通过转换成变分不等式说明供应方的实际受益及成本支出情况，根据各变量调节整体均衡状态并优化活动策略。上述供应方、数字社区和消费者构成的活动服务优化问题是凸函数优化问题，通过变分不等式能够给出有效的活动改善方案。采用变分不等式常用的投影法进行求解，同时获取有关活动服务质量的决策变量和约束条件中拉格朗日乘子的解。

活动优化问题存在各种各样的机器学习算法，一般首选集成学习（Ensemble Learning）方法，该方法结合不同学习器来取得更好的优化性能，尤其在“弱学习”时会取得极佳效果。其中 AdaBoost 算法能够突破分类问题局限性、解决不同场景下的一般决策、回归和分类问题，将其应用在本研究中，效能提升转变为最优规划问题，过程如图 3 所示。

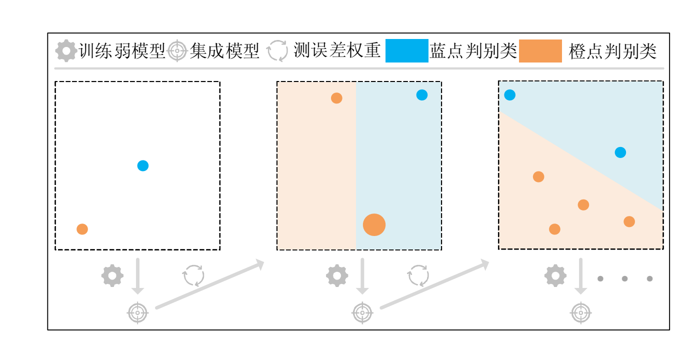


图 3 AdaBoost 效能提升过程示意图

Fig. 3 Schematic diagram of AdaBoost efficiency improvement process

图 3“橙、蓝”两点分别模拟活动三方实际受益和亏损水平，主要用于解释效能提升判别的整体过程，步骤如下：初始化样本数据权重，更新训练样本权重并归一化常数据，按弱分类器权重组合成最终的强分类器，将测试集输入到模型中，并把基学习器作为树桩，当树桩满足阈值标准后通过二叉图实现可视化。对分群结果用户集进行受益和损失判别，从树桩第一个根节点开始进行效能提升，并从全局考虑最终效果。

4 实证分析

在数字中国战略要求下，各传统企业也在不断适应个性化、多样化和不断升级的数字消费趋势，在坚守传统发展历程的同时努力开拓创新，使市场和消费者需求导向得到更多尊重和满足。传统烟草行业在这一方面也做了许多工作，如市场趋势跟踪、消费行为研究、用户精准画像和消费需求挖掘等。基于此，本研究以广西中烟工业有限责任公司为例，对旗下的数字社区消费平台“即开”和“微信社群”数据合并后构造成新的数据集，通过填充缺失值、删除重复数据、量纲标准化和文本预处理等过程，随机选取 5.52 万名用户在 2021 年 1 月—2022 年 7 月的 13.84 万条活动数据用于实证分析。

4.1 用户分群

4.1.1　分群结果

据 4.1 和 4.2 所述方法，利用数据完成画像后通过 DBSCAN 算法对画像进行聚合实现分群，当群数量固定为 6 时，聚类轮廓系数 且群间余弦相似度小于 0.4，表明分群效果良好，结果如表 4 所示。

表 4 表明每个类群在画像、偏好、影响和消费维的具体形态。能够发现，类群 1 主要位于华南地区，从活动时长、兴趣量级、活动及主要倾向折射出该类群具有较强的活动参与度，消费概率和显性消费水平较高，反映出其有较强消费意愿；类群 2 的 UGC 主题为知识传播，且偏好维中相对指数和消费频次较高，

表 4　消费者分群结果

Table 4 Consumer clustering results

所在维度及具体指标		类群 1	类群 2	类群 3	类群 4	类群 5	类群 6
画像维	主要地区	华南地区	华南地区	华北地区	华南地区	华东地区	华南地区
	活动时长	高	中	高	高	中	中
	点赞评论	中	低	中	低	高	高
	UGC 主题	商品宣传	知识传播	活动效果	活动评价	用户意愿	用户祝福
	关键词	龙粉	知识	期待	666	中奖	大卖
	UGC 情感	乐	乐	好	好	惊	乐
偏好维	参与时间	午	午	晚	晚	早	晚
	相对指数	中	高	中	低	高	中
	消费频次	低	高	高	低	中	中
	消费指数	中	低	高	中	中	高
	全局偏好	本体产品	附加产品	附加产品	本体产品	附加产品	本体产品
	兴趣量级	高	中	低	低	高	高
	附加倾向	/	家居	智能	/	旅居	/
	活动倾向	高	中	中	高	低	中
	主要倾向	高	中	高	低	中	低
影响维	用户价值	中	低	高	高	中	中
	裂变 K 值	高	中	低	低	中	高
	盈利得分	中	高	低	低	中	高
消费维	消费层级	中	低	高	高	中	低
	消费概率	高	高	中	低	高	低
	消费能力	中	中	高	高	低	中
	显性水平	高	高	低	中	高	低
	隐性水平	低	高	高	低	低	中

在影响维中的盈利得分和消费维中的隐性消费水平较强，表明该群体多关注活动中的知识流动，并具备一定判断能力和理性消费能力。类群 3 多分布在华北地区，UGC 主题面向活动效果，情感以“好”为主，参与时间多为晚上，且消费频次和指数较高，其活动过程带有一定的娱乐属性，在全局中的附加产品偏好也能对此印证，另一方面该类群用户价值、消费层级和消费能力较强，但裂变 *K* 值和盈利能力相对较低，从侧面反映该类用户粘性较低并具有较强提升空间；类群 4 具有较高的消费层级和能力，但显隐性水平均较低，表明该类用户参与活动时常以“低调”姿态出现，具有重点维护价值；类群 5 的 UGC 关键词为中奖、情感倾向为“惊”，表明活动奖品对其具有较强吸引力；类群 6 的 UGC 主题为“用户祝福”，关键词为“大卖”，表达出这类用户是公司旗下的稳定用户，并对企业抱有较高期待。

4.1.2　命名依据

实现分群后，利用变分法构建数字流量漏斗并对 AARRR 模型进行改写后实现类群命名，根据各自特性，结合调和平均值输出对应分群结果的 6 类群体，分别是：中心类群（约占 17%）、特殊类群（约占 21%）、重点类群（约占 18%）、沉睡类群（约占 5%）、一般用户（约占 30%）和流失类群（约占 9%），这 6 类群体构成数字社区活动用户总和。

4.2　多模态特征分析

4.2.1　共性分析

消费者群间交互作用分析：为判明目标群间交互关系、研究潜在相互作用，根据 4.3 所述方法，将引力、频次阈值和斥力因子分别设置为 0.1、10 和 80 通过共现网络构造交互能力引导图，并基于中心性构造极坐标图，结果表明在消费者群交互网络中活跃度最高和最低的分别是中心类群与沉睡类群，有条件充分运用前者参与数字消费的积极性提升活动效能，对于后者则应及时进行引导和激活，从而挖掘其潜在价值；中心性极坐标图表明不同类群对数字资源的占有程度，其中沉睡类群扇形面积和平均值最小，其资源占据比例最少，中心类群扇形面积和平均值都处于较高水平，表征其活动受益居各类群首位。

消费者群间渗透与漂移现象分析：利用分支定界法识别渗透群体并通过 Relpscost 算法捕获群间漂移现象，结果表明大多数类群都发生有用户渗透现象，较为显著的有重点→一般、特殊→流失、中心→重点、沉睡→流失等群体，能够通过渗透现象挖掘用户群演变机理并针对性提出活动效能提升策略。有一般用户发生了漂移现象，在该类群逐渐消解过程中，仍主要往中心类群方向靠拢，表明前期所得消费者用户的群体成分不仅适用于全局，也能够用来模拟局部演变过程。

4.2.2　特性分析

各个类群都有突出的特点，可以从这些特性入手为服务效能提升策略的部署提供依据。中心类群是数字社区中消费概率和兴趣量级较高且相对稳定的用户成分，应重点分析其沉没成本（用户背离活动所付出的代价），从分群结果能够发现，可以从提升活动消费频次和消费能力出发，增加其沉没成本的同时削弱活动资源占比。重点类群是活动中最具客户价值的用户成分，应对其进行 VIP 服务战略分析和精细运营方案分析，前者可以从分群结果的全局偏好与活动倾向出发来提高其直接感受和体验；后者则从交互关系入手全方位提升用户服务，另外需要避免重点用户向一般转化。一般类群是本研究捕获的唯一发生漂移现象的用户成分，应当尽可能提升其向重点和中心类群漂移的比例，并避免向沉睡和流失类群漂移的倾向。从群间渗透漂移现象来看，特殊类群易转化为流失类群，因此需结合其 UGC 关注主题和关键词对活动参与需求进行重点解读，并从兴趣吸引、产品设置和情感引导入手提升其活动参与意愿。沉睡类群用户同样易转化为流失类群，应从沉寂周期、唤醒缘由和策略出发，避免渗透现象的发生并尽可能加强其与活动关联性。流失类群虽然处在活动的弱关系侧，且占据活动资源较少，但不可忽视的是该类群在消费者成分中约占 13%，首先应该从商业价值评定角度出发分析其流失缘由，然后提出针对性挽回策略。

4.3 服务效能提升的实践对策

利用标准量纲规范化实证数据并将其对应表 3 中的参数、决策变量及函数表，基于三方（供应方、消费者及数字社区）最优行为和均衡条件下的服务效能提升模型构建二叉图，结果详见图 4。

图 4 表明数字消费的最优状态，即从根节点（深度 0，顶部）开始判断用户是否为一般类群，一般类群发展易受各种因素干扰导致失调，产品供应方引导其向非一般类群漂移从而转变为相对可控类群。非一般类群的最优均衡条件是当感知受益时，对活动持续产生较强的情感期待，并尽可能保持消费粘性；当感知亏损时最大程度激活、唤醒处在沉寂观察期的用户。提升树表明用户可以达到的最优均衡状态，基于此给出供应方、消费者及数字社区三方下的服务效能提升路径，结果详见图 5。

图 5 表明对于消费者而言，在活动初期主要考虑活动兴趣和自身参与能力，前者包括对活动时间、回报以及个人偏好因素，后者包括消费者对自我消费能力和水平的评估，能够发现较低成本的数字产品更受青睐；在活动中期，消费者最关心的是活动参与时间和资金上的成本，同时格外关心活动中的实际受益情况；在活动末期，消费者最注重个人活动体验，包括对活动的综合满意度，以及涵盖企业价值观、活动理念等非物质观的认同感。在实践对策方面，可以向用户推荐可能感兴趣的活动或活动关联商品，并降低活动准入门槛，促使用户提高活动参与粘性；应降低用户活动的参与成本，并提高用户感知受益，可以营造生态“社群”氛围、进而为用户创造附加收益；同时，针对用户塑造关于活动的“情怀主义”，或更有助于增加用户活动参与度，直接提升用户对后续活动及企业产品的留存率。

对于数字社区而言，在活动初期应注意营造较强传媒生态，并对活动方案效果进行预判，同时注重数据收集、积累和挖掘，通过解析全方位的消费数据维度建立传媒效果评价模型；活动中期将面临传播方案优化问题，需从推荐内容反馈和传媒内容布局方面改进，并从信息茧房、算法歧视视角，通过机器学习手段为用户过滤可能不感兴趣的内容；活动末期要对整体传播效果进行评估，包括传播内容覆盖率和传播用户准确度两个主要方面，此外还应该对部分高价值数据进行记忆留存，包括用户异常行为数据，中心、重

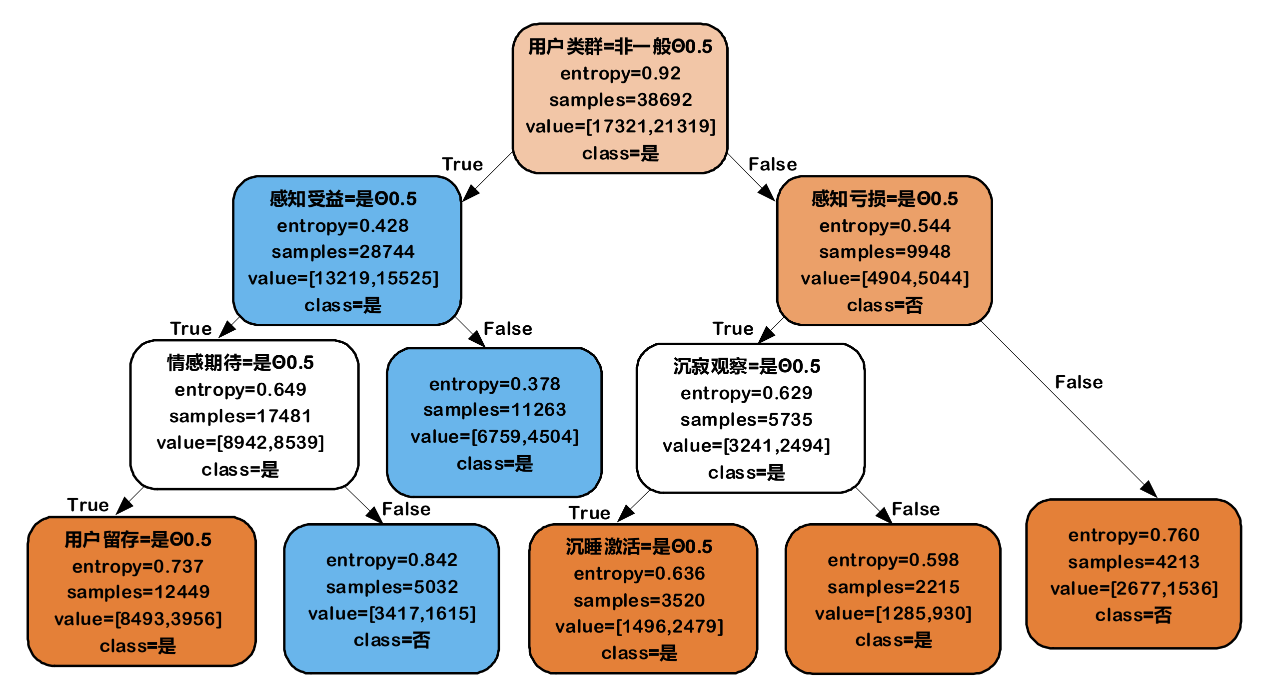


图 4　服务效能提升树下二叉图

Fig.4 Binary diagram under the service efficiency improvement tree

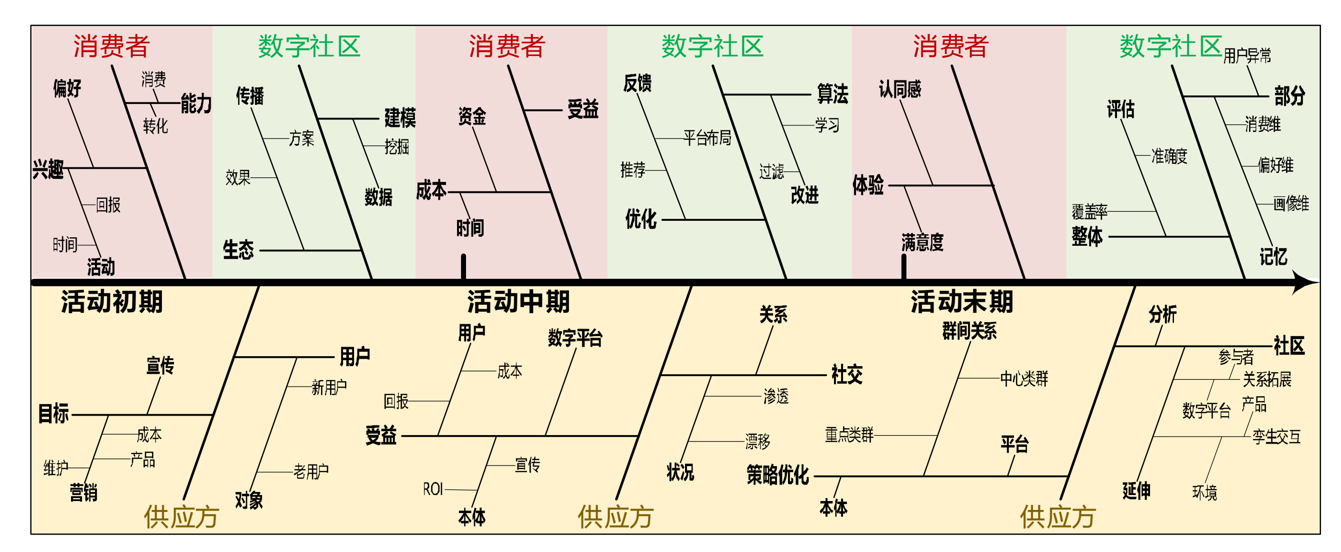


图 5 服务效能提升路径

Fig.5 Service efficiency improvement path

点、特殊、沉睡和流失类群在消费、偏好维以及画像维的数据，以供后续算法优化使用。在实践对策方面，应通过多层嵌套技术对用户数据进行纵深分析，从需求发现角度提取高价值用户；持续优化用户体验，方式包括问卷调查、在线访谈等，根据用户反馈及时改善社区服务水平；同时，开发更完善、适配的活动传播效果评价指标，并不断在实践中修正。

对于供应方而言，在活动初期首先应确定活动目的，包括企业宣传和产品推广以及用户维系等，其次需确定目标受众，针对新老用户特点布局活动方案；在活动中期考虑的受益对象应涵盖企业本体的宣传与 ROI 指标、用户成本与回报、数字社区效益，同时关注目标类群社交关系以及群间渗透和漂移现象，进而驱动用户向三方最优均衡条件转化；在活动末期明确中心和重点类群对企业发展具有较大扶持作用，积极引导其他类群向二者转化。在实践对策方面，可以不断更新产品布局、调整发展态势，将产品供应速度合理放缓、以供服务提升的思路或为可行选择，同时通过关系拓展和数字孪生交互等手段尽可能打造和谐共生、彼此互利的数字生态环境等。

5 结　语

本文在梳理相关研究的基础上，建立数字社区消费者分群特征指标体系，通过 DBSCAN 算法对体系映射的数据进行二次聚合以实现分群，分群后利用 AARRR 模型能够精准识别中心、重点、特殊、沉睡、流失和一般类群用户。绝大多数类群都会产生用户渗透现象，然而仅有一般用户类群会发生漂移现象。实证分析结果表明，本研究方法可以合理有效地实现群类别划分，并针对各个群体特性提供数字社区活动前、中后期的服务效能提升方案。

参考文献：

[1] 邓旻. 以消费者为中心的数字化营销策略研究[J]. 中国市场, 2022(27): 134-136.

DENG M. Research on consumer-centered digital marketing strategy[J]. China market, 2022(27): 134-136.

[2] 中国教育和科研计算机网. “十四五规划”纲要：加快数字化发展 建设数字中国 [EB/OL].[2022-11-02]. https://www.edu.cn/xxh/zt/lhxxh/202103/t20210315_2084688.shtml.

[3] 陈雨倩, 孙虹, 葛王蓉. 数字经济下服装品牌营销策略研究[J]. 经营与管理, 2021(11): 51-55.

CHEN Y Q, SUN H, GE W R. Research on clothing brand marketing strategy in digital economy[J]. Management and administration, 2021(11): 51-55.

[4] 周毅, 白文琳. 公共信息服务政策内容的系统设计[J]. 情报理论与实践, 2013, 36(10): 10-15.

ZHOU Y, BAI W L. System design of public information service policy content[J]. Information studies: Theory & application, 2013, 36(10): 10-15.

[5] 胡振宇. 国内数字营销伦理乱象探因与治理研究——基于数字营销从业精英的访谈[J]. 当代传播, 2018(5): 80-84.
HU Z Y. A study on the causes and governance of ethical chaos in digital marketing in China -Based on an interview with digital marketing professionals[J]. Contemporary communication, 2018(5): 80-84.

[6] LI J, WU Y, XIAO J J. The impact of digital finance on household consumption: Evidence from China[J]. Economic modelling, 2020, 86: 317-326.

[7] 丁盈. "互联网+"背景下农产品数字营销策略研究[J]. 山西农经, 2022(14): 157-159.
DING Y. Research on digital marketing strategy of agricultural products under the background of "internet plus"[J]. Shanxi agricultural economy, 2022(14): 157-159.

[8] 郭褕亦. 泛媒体时代中文播客的数字营销——以小宇宙 App 为例[J]. 新媒体研究, 2021, 7(17): 47-49.
GUO C Y. Digital marketing of Chinese podcasts in pan-media era - A case study of small universe App[J]. New media research, 2021, 7(17): 47-49.

[9] 鲍磊, 江梓毓, 奚凯悦. 数字营销的发展：从 1.0 时代到 4.0 时代的进化[J]. 山东纺织经济, 2022, 39(2): 19-22.
BAO L, JIANG Z Y, XI K Y. Development of digital marketing: Evolution from era 1.0 to era 4.0[J]. Shandong textile economy, 2022, 39(2): 19-22.

[10] 高培培. VR 技术下数字营销专业虚拟仿真实训中心的建设与实践[J]. 营销界, 2022(1): 65-67.
GAO P P. Construction and practice of virtual simulation training center for digital marketing specialty under VR technology[J]. Marketing circles, 2022(1): 65-67.

[11] 苗月新. 基于营销宏观环境理论的网上零售影响因素分析[J]. 中国市场, 2022(5): 125-127.
MIAO Y X. Analysis on influencing factors of online retail based on marketing macro-environment theory[J]. China market, 2022(5): 125-127.

[12] DONG Y F, TAN R H, ZHANG P, et al. Product redesign using functional backtrack with digital twin[J]. Advanced engineering informatics, 2021, 49(3): 101.

[13] YOON S, MCCLEAN S T, CHAWLA N, et al. Working through an "infodemic": The impact of COVID-19 news consumption on employee uncertainty and work behaviors[J]. Journal of applied psychology, 2021, 106(4): 501-517.

[14] CUMMINGS B. Digital assets add new concerns to estate planning[J]. Journal of financial planning, 2022, 35(8): 19.

[15] 徐婷. 基于新技术的数字营销战略——以 AllSaints 为例[J]. 内蒙古煤炭经济, 2020(22): 87-89.
XU T. Digital marketing strategy based on new technology - Taking AllSaints as an example[J]. Inner Mongolia coal economy, 2020(22): 87-89.

[16] KAPKAEVA N, GURZHIY A, MAYDANOVA S, et al. Digital platform for maritime port ecosystem: Port of hamburg case[J]. Transportation research procedia, 2021, 54: 909-917.

[17] 王佳. 数字化时代商业银行优化数字营销方式策略[J]. 中国市场, 2021(29): 119-120.
WANG J. Strategies for commercial banks to optimize digital marketing mode in the digital age[J]. China market, 2021(29): 119-120.

[18] 邓溱薇. 数字营销人才能力模型构建研究[D]. 杭州: 浙江工商大学, 2022.
DENG Z W. Research on the construction of competence model of digital marketing talent[D]. Hangzhou: Zhejiang Gongshang University, 2022.

[19] FUCHS C. The digital commons and the digital public sphere how to advance digital democracy today[J]. Westminster papers in communication and culture, 2021, 16(1): 9-26.

[20] SKARE M, SORIANO D. How globalization is changing digital technology adoption: An international perspective[J]. Journal of innovation & knowledge, 2021, 6(4): 222-233.

[21] 王宪朋. 基于视频大数据的用户画像构建[J]. 电视技术, 2017, 41(6): 20-23.
WANG X P. Construction of user portrait based on video big data[J]. Video engineering, 2017, 41(6): 20-23.

[22] 王帅. 突发公共卫生事件情境下在线健康社区用户画像与分群

研究[J]. 情报科学, 2022, 40(6): 98-107.

WANG S. Study on user portrait and clustering of online health community in the context of public health emergencies[J]. Information science, 2022, 40(6): 98-107.

[23] 李旭光, 肖思琪, 李珊珊, 等 基于知识行为的小米社区用户画像研究[J]. 农业图书情报学报, 2021, 33(8): 4-12.

LI X G, XIAO S Q, LI S S, et al. Research on user profiles of Xiaomi community based on knowledge behavior[J]. Journal of library and information science in agriculture, 2021, 33(8): 4-12.

[24] 曹树金, 岳文玉. 面向精准服务的图书馆用户画像研究[J]. 农业图书情报学报, 2021, 33(10): 4-19.

CAO S J, YUE W Y. Research on library user profiles for precision services[J]. Journal of library and information science in agriculture, 2021, 33(10): 4-19.

[25] LI D. Text mining model for virtual community user portrait based on social network analysis[J]. Tehnicki vjesnik - Technical gazette, 2019, 26(4): 1145-1151.

[26] MIN T, CAI W X. Portrait of decentralized application users: An overview based on large-scale Ethereum data[J]. CCF transactions on pervasive computing and interaction, 2022, 4: 124-141.

[27] 刘海鸥, 孙晶晶, 苏妍嫄, 等. 国内外用户画像研究综述 [J]. 情报理论与实践, 2018, 41(11): 155-160.

LIU H O, SUN J J, SU Y Y, et al. Literature review of persona at home and abroad[J]. Information studies: Theory & application, 2018, 41(11): 155-160.

[28] 简宋全, 李青海, 秦于钦. 基于 K-means 的用户分群分析[J]. 现代计算机(专业版), 2017(29): 29-31.

JIAN S Q, LI Q H, QIN Y Q. User group analysis based on K-means[J]. Modern computer, 2017(29): 29-31.

[29] 占张帆. 基于用户分群的产品设计知识推送研究[D]. 南昌: 南昌航空大学, 2016.

ZHAN Z F. Research on product design knowledge push based on the user group[D]. Nanchang: Nanchang Hangkong University, 2016.

[30] 宋嵩. 基于 URL 分析的移动互联网用户分群 [D]. 保定: 河北大学, 2013.

SONG S. Mobile Internet user clustering based on the analysis of URL[D]. Baoding: Hebei University, 2013.

[31] 吴艳玲, 孙思阳. 基于马尔可夫模型的图书馆用户聚类分群方法研究[J]. 情报科学, 2021, 39(11): 167-172.

WU Y L, SUN S Y. Clustering method of library users based on Markov model[J]. Information science, 2021, 39(11): 167-172.

[32] 王珧. 大数据时代网络营销策略分析[J]. 海峡科技与产业, 2015 (12): 94-96.

WANG Y. Analysis of network marketing strategy in the era of big data[J]. Technology and industry across the straits, 2015(12): 94-96.

[33] 陈长庆. 在线旅行社客户关系管理研究——以途牛网为例[D]. 南京: 南京农业大学, 2017.

CHEN C Q. Research on customer relationship management of online travel agency - With Tuniu network as example[D]. Nanjing: Nanjing Agricultural University, 2017.

[34] 王婉, 张向先, 郭顺利, 等. 基于扎根理论的移动专业虚拟社区用户需求模型构建[J]. 情报科学, 2022, 40(6): 169-176.

WANG W, ZHANG X X, GUO S L, et al. Construction of user demand model of mobile professional virtual community based on grounded theory[J]. Information science, 2022, 40(6): 169-176.

[35] 张邦辉, 万秋兰, 吴健. 在线政务服务的营商环境优化效应探析——"数字红利"与"数字鸿沟"[J]. 中国行政管理, 2021(4): 70-75.

ZHANG B H, WAN Q L, WU J. Analysis of optimizing business climate by online government services: "Digital dividend" and "digital gap" effects[J]. Chinese public administration, 2021(4): 70-75.

[36] 于丽娟. "互联网+智慧服务"优化数字阅读精准推广模式研究[J]. 河南图书馆学刊, 2021, 41(9): 4-6.

YU L J. Research on "Internet+smart service" optimizing digital reading accurate promotion model[J]. The library journal of Henan, 2021, 41(9): 4-6.

[37] 王毅, 吴睿青. 公共图书馆数字文化资源服务用户画像研究[J]. 图书情报工作, 2021, 65(16): 42-55.

WANG Y, WU R Q. Research on user portrait of digital cultural resource service in public library [J]. Library and information service, 2021, 65(16): 42-55.

[38] CHEN Y B, HE J S, WEI W, et al. A multi-model approach for user portrait[J]. Future Internet, 2021, 13(6): 147.

[39] SUN Y, CHAI R Q. An early-warning model for online learners

based on user portrait [J]. Ingénierie des systèmes d´information, 2020, 25(4): 535-541.

[40] WANG X, WEI X, MA J, et al. User portrait technology and its application scenario analysis [C]//BDE 2021: The 2021 3rd international conference on big data engineering, New York: Association for Computing Machinery, 2021: 64-69.

[41] 石光莲, 张敏, 郑伟伟. 形式概念分析在 Folksonomy 中的应用研究进展[J]. 图书情报工作, 2014, 58(9): 136-142.

SHI G L, ZHANG M, ZHENG W W. Advances in applied research of formal concept analysis in folksonomy[J]. Library and information service, 2014, 58(9): 136-142.

[42] 张淑萍. 长三角城市群国际消费中心城市竞争力评价[J]. 商业经济研究, 2022(10): 44-49.

ZHANG S P. An evaluation of international consumption center cities' competitiveness in Yangtze River Delta urban agglomeration[J]. Journal of commercial economics, 2022(10): 44-49.

[43] 王晖, 王晋立. 北京居民夜间消费倾向实证研究——基于Ⅳ-logit 回归模型的分析[J]. 城市问题, 2021(2): 75-83.

WANG H, WANG J L. The empirical study on night consumption tendency of Beijing residents: Based on Ⅳ-logit regression model[J]. Urban problems, 2021(2): 75-83.

[44] 刘启林. 用户画像的基础、原理、方法论(模型)和应用[EB/OL]. [2022-11-06]. https://zhuanlan.zhihu.com/p/140104236.

LIU Q L. The basis, principle, methodology (model) and application of user portrait[EB/OL]. [2022-11-06]. https://zhuanlan.zhihu.com/p/140104236.

[45] LIN Z, CHAO M G. Research on acquisition on authors sources of sci-tech journals based on AARRR model [C]//ICCIR'21: Proceedings of the 2021 1st international conference on control and intelligent robotics, New York: Association for Computing Machinery, 2021: 193-196.

[46] 詹娟娟. 概率无向图模型近邻传播聚类算法的研究[D]. 南宁: 广西大学, 2017.

ZHAN J J. Research on affinity propagation clustering algorithm for probabilistic undirected graph model[D]. Nanning: Guangxi University, 2017.

[47] 李沛, 顾东蕾, 邹涛, 等. 基于社会网络分析的本科生社会科学潜在阅读兴趣小组挖掘探析——以中国药科大学为例[J]. 农业图书情报学报, 2021, 33(3): 78-89.

LI P, GU D L, ZOU T, et al. Exploration of potential reading interest groups in social sciences of undergraduates based on social network analysis: A case study of China pharmaceutical university[J]. Journal of library and information science in agriculture, 2021, 33(3): 78-89.

[48] 胡昌平, 胡吉明, 邓胜利. 基于 Web2.0 的用户群体交互分析及其服务拓展研究[J]. 中国图书馆学报, 2009, 35(5): 99-106.

HU C P, HU J M, DENG S L. Analysis of network users' group interaction and research for service based on the web 2.0[J]. Journal of library science in China, 2009, 35(5): 99-106.

[49] GASSE M, CHETELAT D, FERRONI N, et al. Exact combinatorial optimization with graph convolutional neural networks[C]//Advances in neural information processing systems 32 (NeurIPS 2019), Vancouver Canada: MIT Press, 2019: 15580-15592.

[50] 孟秀丽, 吴一凡, 刘波. 考虑延误险的多期众包物流服务质量优化[J/OL]. 中国管理科学: 1-15[2022-11-08]. DOI:10.16381/j.cnki.issn1003-207x.2021.1807.

MENG X L, WU Y F, LIU B. Multi-phase crowdsourcing logistics service quality optimization considering delay insurance [J/OL]. China management science: 1-15[2022-11-08]. DOI:10.16381/j.cnki.issn1003-207x.2021.1807.

[51] NAGURNEY A, DANIELE P, SHUKLA S. A supply chain network game theory model of cybersecurity investments with nonlinear budget constraints[J]. Annals of operations research, 2017, 248(1/2): 405-427.

Multi-modal Characteristics Analysis and Customer Service Efficiency Improvement in the Digital Community Based on User Clustering

LI Canyao[1], WEI Wei[1], LIU Xiaoli[1], ZHOU Linxing[2*], WANG Shuai[2]

(1. Guangxi China Tobacco Industry Co., Ltd., Nanning 530001; 2. School of Cultural Heritage and Information Management, Shanghai University, Shanghai 200444)

Abstract: [Purpose/Significance] Multi-modal feature analysis and service efficiency improvement of digital community consumers will help to provide a new vision for the construction of digital intelligent online communities and provide new impetus for relevant departments to make decisions. In addition, although the current research on digital consumption includes the relevant content of user analysis, it mainly aims at the formulation of detailed operation plans, and lacks the analysis of service efficiency improvement of digital communities. On the other hand, the research on user value orientation for online service quality optimization is mostly based on profile technology, which only considers the difference characteristics of a single target user, and lacks the horizontal comparison and difference attribution research of multi-modal features among groups. Based on this, this paper, from the perspective of value discovery, achieves clustering by aggregating user profiles, analyzes the multi-modal characteristics of consumer groups in digital communities, and proposes a service efficiency improvement scheme. [Method/process] First, this paper analyzed the target consumers in the digital community and established a cluster indicator system. Then, users were grouped, and the multi-modal information profile of the target group was restored based on group characteristics and inter-group interaction characteristics. Finally, it proposed the path to improve the efficiency of digital community services. In terms of technical implementation, the data related to consumer activities were extracted from the digital community, integrated, cleaned, and distributed to the storage bucket. The clustering indicator system was built through feature mining and existing indicators, and the indicators were mapped to aims, and DBSCAN clustering was carried out on the basis of using AP to realize the image. After grouping and naming, the characteristics analysis, interaction analysis, and drift and penetration phenomenon analysis were carried out according to the characteristics of various groups. We extracted various parameters of the design of digital community consumption activities, and built a decision variable function to find the optimal behavior equilibrium conditions of the digital product supplier, consumer and digital community. Based on this, we built an efficiency improvement tree, and proposed community service efficiency improvement strategies at the initial, middle and later stages of consumption activities. [Results/Conclusions] The empirical analysis results show that the model in this paper can first generate reasonable and effective clustering results, and then realize the classification of group characteristics and the analysis of inter-group infiltration and drift. The clustering results show six types of consumer groups: focus, center, special, sleeping, loss and general groups. Most groups will have user penetration, and only general user groups will have inter-group drift. The service efficiency improvement model shows that the most valued group is the center and key group. The inadequacy of this study is that the applicability of the model to multi-source heterogeneous data needs to be tested and there is still room for improvement in clustering granularity.

Keywords: user clustering; AP-DBSCAN; multi-modal characteristics; digital community; digital consumption